Biol Philos (2017) 32:887–904
https://doi.org/10.1007/s10539-018-9610-x

The evolution of linguistic rules

Matthew Spike[1]

Received: 23 June 2017 / Accepted: 20 February 2018 / Published online: 26 February 2018

Abstract Rule-like behaviour is found throughout human language, provoking a number of apparently conflicting explanations. This paper frames the topic in terms of Tinbergen's four questions and works within the context of rule-like behaviour seen both in nature and the non-linguistic domain in humans. I argue for a minimal account of linguistic rules which relies on powerful domain-general cognition, has a communicative function allowing for multiple engineering solutions, and evolves mainly culturally, while leaving the door open for some genetic adaptation in the form of learning biases.

Keywords Language · Linguistic rules · Evolution · Cognition · Tinbergen · Learning biases · Domain specificity

Introduction

The Cambridge Grammar of the English Language, "a detailed account of the principles governing the construction of English words, phrases, clauses, and sentences" (Huddlestone and Pullum 2002, p. 2), is 1,860 pages long. Introductory texts covering the linguistic features of any language run into the hundreds of pages. These works are not simply lists of arbitrary linguistic conventions.[1] Although human languages are shot through with irregularities, linguists are typically drawn towards their ubiquitous regularities and rule-like structures. This is reflected in the vast

[1] cf. Strunk and White (1979).

✉ Matthew Spike
matthew.spike@anu.edu.au

[1] Centre of Excellence for the Dynamics of Language, Australian National University, Coombs Building 9, Room 2220, Canberra, ACT 2612, Australia

literature on linguistic theory.[2] These theories are sometimes framed—where they are framed at all—in terms of descriptive accuracy, with no recourse to psychological or evolutionary plausibility. Elsewhere, we see the opposite: Pinker (1991, 1999) has argued that rule-like behaviour is the defining characteristic of human language, and that these rules are innate, domain-specific, and evolved via natural selection under communicative pressures. I will argue mostly against this view, in light of work on human and comparative cognitive evolution. Linguistic rules involve sophisticated but *general* capacities for cognitive processing; are socially learned; have a general but not specific communicative function; and arise via analogical reasoning in a cultural evolutionary process. However, the sheer ubiquity and variation of rule-like structures in human language demands some explanation: mine is that language evolves (mainly) culturally under pressure to reliably maintain communicative function, where multiple functionally-equivalent solutions are possible.

This paper will proceed as follows: I first look at what form linguistic rules can take. I then use the framework provided by Tinbergen (1963) to survey different types of explanations for the existence of linguistic rules. Next, I consider this alongside knowledge about related human and non-human capacities, and use this to propose a story which makes the minimum of theoretical assumptions, namely that rules are a mostly culturally-evolved, learned phenomenon with an overall but underspecified communicative role, but leaving the door open for a somewhat diminished role for biological adaptation and domain-specificity.

Finally, as a caveat, the scope of this article does not include the evolution of natural language syntax, let alone recursion, hierarchical structure, or the upper echelons of the Chomsky hierarchy. One more thing: it is not common to treat 'linguistic rules' in generality, without sticking to a specific linguistic domain, such as phonology or morphology. On the other hand, researchers within these sub-disciplines make a number of rather strong claims regarding mechanism, function, acquisition, and evolution which I feel are best treated at a more general level. As such, my motivating question is: what are linguistic rules, why are they so ubiquitous, and why do they share some features while also displaying such variation?

What are linguistic rules?

Here are some rule-like behaviours found in many varieties of English. They are written out in a standard linguistic formalism, but are explained afterwards.

(1) *ŋ/ #_
(2) t → [ɾ]V_V
(3) Evaluative > General property > Age > Colour > Provenance > Manufacture > Type

[2] As an example, the Encyclopaedia of Language and Linguistics (Brown 2005) conservatively lists 13 distinct schools of phonology, 12 models of morphology, and 27 syntactic frameworks.

Example (1) is a *phonotactic constraint*, and states that the sound /ŋ/ (the final sound in the words "bang" or "flung") never occurs word-initially in English. Example (2) is a *phonological rule* referred to as 'intervocalic flapping', seen for example in many American and Australian English pronunciations of the words 'batter' or 'butter': the unvoiced /t/ is pronounced as the more 'd-like' flap /ɾ/ when situated between two vowels. Finally, Example (3) is a description of the order of noun modifiers in English,[3] seen in the example "those silly little old red Hawaiian beach shorts".

An important fact about these rules, and other rules like them, is that they are not only specific to English, or at least certain varieties of the English, but often exist in variation even within those varieties. For example, while Vietnamese or Warlpiri words freely break Example 1, no dialects of English do so. However, some English dialects display nothing like Example 2; others do see a similar change, but to a different consonant altogether (e.g. to the glottal stop /ʔ/ in many varieties of UK English). As for Example 3, as Huddlestone and Pullum (2002) point out, it is easy to find contradictions to any one of the orderings (e.g. 'big bad wolf'), often depending on the linguistic context, and irrespective of dialect.

This kind of behaviour is found across all levels of linguistic analysis. For example: (1) *prosody*: in English, a fall-rise tone at the end of a sentence signals a question, questions in Finnish exhibit the opposite rise-fall pattern; (2) *morpho-phonology*: e.g. pluralisation in English involves three suffixes, /s/, /z/, and /ɪz/, selected based on the preceding sounds: similar processes are found in all languages with complex words; (3) *morphology*: processes such as agreement for person, number, and gender, and restrictions on the concatenative order of suffixes, prefixes, and infixes are extremely prevalent cross-linguistically; (4) *serial word order*: demonstratives (e.g. 'this' and 'those') occur before adjectives in English, but after them in Turkish. Moreover, languages exhibit a multitude of rule-like behaviours of types entirely absent from English, such as Turkish vowel harmony, or the scaffold-like semantic patterning of the Semitic root system. The point is that such regularities found in language are many, varied, and—at first analysis—seemingly arbitrary.

Where do rules come from?

The flapping rule in Example 1 above seems well-motivated purely on the basis of human anatomy. The unvoiced consonant, when surrounded by vowels, becomes more vowel-like itself; a similar explanation could be made for the rules governing pluralisation in English. Some linguists (e.g. Blevins 2004; Ohala 2005) think that the explanation for most rule-like behaviour, in phonology at least, is along these lines. They appeal to 'phonetic naturalness': physical limitations on the way we produce and interpret sound, established and compounded over time, lead to the kind of rule-like process described above. This kind of approach predicts the *diachronic* (i.e. over time) emergence of regular rule-like behaviour, without specifying

[3] What Huddlestone and Pullum (2002) call the 'labile ordering of residual pre-head modifiers'.

exactly where and how it will form. In terms of Example 2, it would say that (1) the inter-vocalic environment is a likely place for some kind of regular sound change to emerge, and (2) that if it does, it is likely to take to have the form 'consonant becomes more vowel-like'. As such, although the consequences of phonetic naturalness are under-specified, they can lead to the emergence of deterministic rules. Other phonologists (e.g. Anderson 1981; Hale and Reiss 2008; de Lacy and Kingston 2013) disagree with these 'naturalistic' accounts, and propose that rule-like behaviour is essentially *synchronic* in nature, i.e. the result of an active cognitive process.

This typifies a fundamental disagreement found in linguistics. It regards not just cognitive sophistication and its relation to cultural aspects of language, but also the importance of communication and learning. However, the exact nature of the disagreement is not always readily apparent, as the linguists involved often have quite different explanatory targets. For 'naturalists', rules are an emergent phenomenon, contingent on history alone, so they focus on historical process. 'Cognitivists' see rule-like behaviour as a 'surface' phenomenon derived from underlying sets of interacting representations, rules, or constraints, so these become the target of their investigation. Despite this, neither account can dispense with some kind of synchronic/cognitive or diachronic/cultural explanation; rather, it is often left implicit. For example, in explaining why American English exhibits flapping while most British varieties do not, a naturalist would outline a process shaped by physical function and blind historical contingency: it is easier to articulate a /t/ as either an /ɾ/ or a /ʔ/ when between vowels, but the former occurred and fixated in American English, the latter in some varieties of UK English, and in still other dialects nothing happened. A cognitivist account,[4] on the other hand, might propose that *all* speakers have a representation of several rules, with American speakers privileging the /ɾ/ rule, UK speakers the /ʔ/ rule, and so on. Neither would deny some role for history or cognition, but would likely identify one or the other as where the 'real action' was happening. To resolve this dispute, we might look to the fact that other languages exhibit very similar flapping, but naturalists would again point to historical and functional convergence, cognitivists to a shared, internal apparatus. Or we might turn to other types of rules, such as Example 3 above: is the serial order of nominal modifiers driven by some covert cognitive mechanism? After all, *most* productions in English follow this non-obvious rule most of the time. Or is it purely convention, as suggested by the fact that other languages choose quite different strategies, and even English will happily break the rule when contextually convenient? Once again, the answer depends on whether the focus is on the cognitive or cultural implications of the rule.

At first glance, it might seem that the burden of proof is on more cognitive explanations: why posit more complex mental structures than are necessary, as long as humans can represent the 'surface' rules? In response, many linguists argue that the best place to find cross-linguistic similarities is not in surface phenomena, or in their functional constraints, but at a more abstract, cognitive level (in accord with their theories). More than this, the presence or lack of an apparent functional explanation

[4] In this case, something vaguely resembling Optimality Phonology.

is often taken to imply whether a phenomenon is a purely cultural or cognitive one. For an example, we can turn again to the idea of naturalness in phonology, and the existence of 'crazy' or 'unnatural' rules. Much has been made of the fact that there is no apparent phonetic or physiological basis for many rules across all languages. Many (e.g. Buckley 2000) see this as clear evidence that their theory not only describes psychological reality, but also an innate linguistic module. Citing the same evidence, however, Blevins (2008) argues that the unnaturalness of any rule is arguable at best and unprovable at worst, that phonetically-unmotivated grammars are only descriptively useful, and that no cognitive module is even required. It is hard to find a linguistic case in which functional—i.e., physiological—concerns are more overt, and hence implications for cognition and culture, but even here the debate is fundamentally coloured by theoretical predisposition.

These positions seem hard to reconcile. Johansson (2013) suggests a way to resolve such thorny issues in linguistics: treat them to an analysis in the spirit of Tinbergen (1963). The following section re-examines the debate using this framework.

Tinbergenian explanations for linguistic rules

Tinbergen (1963) outlined four elements required of any complete biological explanation: (1) 'survival value', i.e. function, (2) 'ontogeny', i.e. acquisition and development, (3) 'evolution', i.e. phylogenetic history, and (4) 'causation', i.e. mechanism. Tinbergen's stated aim was to clear up a perennial debate in his field, ethology, regarding what "science is about," but his analysis extends to biology as a whole. Crucially, the 'four questions' are meant to be not in competition, but instead complementary, moving away from disagreement about what constitutes the 'best kind' of explanation.

The study of human language is in a similar position. As has been seen above, some researchers are drawn to functional explanations, others mechanistic, and so on. This is fine, but when debate turns to the *priority* of different kinds of explanation, Tinbergen's framework applies. This is especially the case when focussing one explanatory aspect can lead to shaky assumptions about another. As an example of this kind of cross-talk, look at how some of the accounts above play out. For the cognitivists, the logic seems to be: we have deduced a rich, nuanced mechanism, but there is either no readily apparent function beyond communication (e.g. Pinker and Bloom 1990) or no function or relation to ontogeny at all (e.g. Buckley 2000): therefore it must be innate, therefore evolution/historical processes are largely irrelevant. The naturalists mirror this: we have deduced a rich, nuanced historical (cultural evolutionary) account, which is driven by function and ontogeny (e.g. Ohala 2005; Blevins 2008): therefore there is no role for innateness, and mechanism is largely irrelevant. Not only is there no a priori reason why either line of reasoning should be preferred, there are many alternatives to each: for example, both cognitive and naturalistic accounts can be framed as a process of gene-culture co-evolution, e.g. Christiansen et al. (2011) and Thompson et al. (2016), in which *all* four questions necessarily apply. In order to disentangle these different concerns, the following sections

map out the landscape of how linguistic rules have been explained previously, but explicitly in terms of Tinbergen's four questions.

The mechanism of rules

Language obviously involves *some* physical and mental mechanisms. What has proven to be notoriously controversial—both inside and outside traditional linguistics—is the nature, sophistication, and explanatory scope of these mechanisms. Putting aside debates about what constitutes the 'real' study of human language, any mechanistic story needs to provide an answer to at least two questions. The first is how rule-like behaviour integrates with the observable use of language—and this must be addressed even if the answer reached is its irrelevance. More specifically, there must be some treatment of the fact that language production involves finely-tuned top-down articulatory control, irrespective of modality, and that language comprehension involves the rapid decoding of a transient sequential input. On the other hand, the second question requires that the external realities of language use should be connected to some cognitive story, however reduced. As such, we want to know the formal shape of such rules, how they are mentally represented, and how they interact with other aspects of cognition. This is no mean feat, and the range of explanations found in the literature is predictably wide in both regards.

Many (often functionally-oriented) linguists stress the sophistication of language use and the complexity of human social interaction. How does this bear on the mechanism of linguistic rules? Blevins (2008, p. 141), arguing for an entirely naturalistic analysis of phonology, cites the "growing literature on language as a complex self-organizing system, with positive and negative feedback loops, and multiple interacting levels." Notably, these levels are *emergent*: they arise via a continuous process of analogical re-analysis during the repeated and messy processes of production and comprehension. As such, they require little in the way of dedicated cognitive machinery. For example, although something like the phoneme /ə/ clearly exists in the context of English speech, it can't be said to map onto any distinguishable mental object, and the same would be the case for a combinatorial phonological rule (e.g. Oudeyer 2005; Zuidema and de Boer 2009). Mechanisms focus on production and comprehension, and in the case of error, analogical repair. Rules take shape via the gradual compounding of error and analogy, and spread throughout the linguistic system. The generalisations involved in these rules are cultural accidents which may fade or change over time, but are not taken to represent any underlying cognitive architecture. What, then, is the cognitive architecture required in this case? An ability to (1) store a large number of 'fixed sequences' which display the *semblance* of rule-like behaviour, then (2) generalise a rule from the reasonably rich, often physiologically salient data, and finally (3) analogically apply that rule to a novel context, possibly during a process of largely habitualised yet still complex production or comprehension. This kind of largely holistic, mass-storage take on language is found in several linguistic domains, but is particularly strongly evidenced in exemplar- and construction-based approaches (e.g. Croft 2001; Walsh et al. 2010). However, although it is supposedly less 'cognitive', it can hardly be said

to have removed cognitive demands completely, or even at all: it requires a large memory, near-instantaneous recall, and the ability to reason analogically on the fly [see Christiansen and Chater (2016) for a detailed account of what this kind of processing-based story might look like]. What it does do, though, is remove the need for any structured, language-specific process.

At the other end of the spectrum, cognition-heavy theories about the nature of linguistic rules abound. When Pinker (1999) refers to 'digital minds in an analogue world', our minds—and especially language—are taken to be fundamentally algorithmic in nature. This is very much in line with much mainstream thought in generative linguistics [see Boeckx (2010)], and the many rule-based approaches along those lines. From this perspective, it is easy to see rule-like structures in language as simply the external manifestation of internal process, although possibly modified or transformed in some way. However, this is not the position taken by all generative linguists: Berwick and Chomsky (2015) discuss the complexities of processing and use, and argue that any *non-syntactic* rule-like behaviour observed in the process of what they term 'linguistic externalisation' is purely that—external—and cite parallels with the songs of birds and cetaceans: they are surprisingly sympathetic to naturalistic explanation, and to the influence of general constraints of processing and use on this external form. Many other accounts, however, do assume a richer mentalistic explanation for rule-like behaviour (e.g. Anderson 1981; Hale and Reiss 2008). In these accounts, dedicated cognitive systems govern what is and isn't possible within some domain, for example phonology or morphology: these internal representations are then filtered through into externalised language. Much of the debate within these circles centres around unearthing the most parsimonious theory to account for the linguistic data, often taken (at least implicitly) to be representative of actual cognitive processes. While early work such as that of Chomsky and Halle (1968) led to an explosion of apparently arbitrary rules, later theories found ways to abstract and generalise, whether these were in the form of innate structural principles (e.g. autosegmental phonology, government phonology), or more psychologically-inspired systems of ranked, but presumably universal constraints (e.g. optimality theory).

Turning briefly to cognitive accounts at other levels of analysis (along the lines of Marr and Poggio 1976) , the relative sophistication of non-algorithmic accounts can be hard to evaluate. 'Implementational' level descriptions such as the neural-network models of Elman (1991) demonstrate that rule-like behaviour can be learned and produced by frameworks which do not themselves look particularly algorithmic, and are rather general in nature; 'computational' level accounts (usually some flavour of Bayesian, e.g. Griffiths et al. 2010) are able to eschew the details of process altogether, and relate input to output in terms of one of many functions over a probability distribution. Two things are worth noting here. Firstly, while both neural networks and computational mechanisms are almost always stochastic in some regard, algorithmic accounts of language traditionally were not. However, this is by no means necessarily true: stochastic algorithms have proven very successful and are used extensively in applied, corpus-based, and computational linguistics. Secondly, there is a long-standing debate about the complementarity of the different levels of analysis, Marr's or otherwise (e.g. Griffiths et al. 2015). Nevertheless, despite the difficulty in ascertaining the correct level of analysis, it is quite reasonable to

assume they all make similar *computational* demands, for example on processing and memory. In all cases, we must assume that processing is fast and sophisticated, and memory is large (although how large depends on theory).

To sum up, what we see in the literature is lot of work aiming to either minimise or maximise the cognitive burden *dedicated* to linguistic rules. At the same time, the significant articulatory and processing demands of actual use are assumed by either side: as *explanatory* for more naturalist/functionalist adherents, and *irrelevant* for most (but not all) of the cognitively-oriented. In fact, given that both sides of the debate require reasonably powerful mechanisms, the only real difference appears to be regarding their domain-specificity.

The function of rules

Identifying the function of any evolved system is difficult (see Gould and Lewontin 1979), and linguistic rules are no exception. Even then, if we can identify some signature of function, we then need to allocate it between communicative, social, and broader cognitive and physiological concerns. More than this, some rules are *functionally structured*, while others are not: more on this later. Finally, and further muddying the waters, this is an area where individuals happily take a strong functional position for one linguistic domain, and oppose it absolutely for another. Tinbergen's definition of function is helpful here: by concentrating explicitly on the *synchronic* function of linguistic rules, we can evade some of these issues. In fact, removing the *historical* aspect means that we can uncontroversially identify a global, communicative/interactive function for most rules, but only sometimes identify a function specific to the shape of that rule.

Language is used for communication and social interaction, so it might seem redundant to claim this is as its function. However, some—notably Berwick and Chomsky (2015)—would certainly not say this. This is because they have a very specific idea about what constitutes human language: it applies only to syntax, and almost any other recognisable feature of language is relegated to the status of either 'externalisation' or 'internalisation'. For both of these—and this includes all aspects of language use, as well as phonology and any non-syntactic phenomenon, including semantics—Berwick and Chomsky (2015) are content to align themselves with the mainstream: language is *used* for communication: they would just deny that this is its primary function, or a factor in its evolutionary emergence. This being the case, and given that we're staying away from syntax, most would agree that linguistic rules are externally employed in and for communication and social interaction. However, this is a *global* property of language: its context of use is *almost always* communicative and social. We can ask a more specific question: is there something about the *structure* of a particular rule which has function?

Rules have structure, but this structure is only sometimes functional in of itself. As a non-linguistic example, consider food preparation: catch a goat → butcher the goat → boil the goat → serve the goat → eat the goat. In this case, the individual actions have a clear purpose, as does the sequence (getting it wrong would lead to strange or unsanitary outcomes), as does the overall task, i.e. providing food. But

this is not necessarily the case: if I add several spices to the boiling goat, the order is probably irrelevant in terms of final outcome, but it is quite possible that I do it the same way every time. This might simply be how I learned to do it, or how the spices are arranged on a shelf, but these are purely historical or incidental factors which have no effect on actual function. Goat preparation is *functionally structured*, but spice addition is not clearly so. And this is not the end of the story: I might stir each spice in individually, to avoid clumping. Now function is clear, but it is entirely structural: the individual elements and their sequence is irrelevant. Adding spices has an *interchangeable* quality, goat preparation a *relational* one.

Moving back to language, we can first identify (1) some rules without functional structure, i.e. which have function only by virtue of being involved in language use. Rules which do have functional structure can be subdivided into (2) interchangeable rules, where simply the presence of an arbitrary structure has function, and (3) relational rules where the nature of composition is significant. The 'crazy rules' discussed above are a good example of the first category: while debate has centred around whether they they have a good historical explanation, their apparent 'unnaturalness' implies that their impact on communication, for example, is either neutral or negative. Turning to an interchangeable rule, all languages have a prescribed syllable structure, but this structure varies greatly. Hawaiian has almost no consonant clusters; in Georgian, six or more consonants are often strung together. However, no language allows an arbitrarily long sequence of consonants, for which the communicative benefits seem clear: Pinker and Jackendoff (2005) argue along these lines for phonology as supporting a 'parity principle', in which a phonological code is optimised for use by both speaker and hearer, but where the specific code is somewhat arbitrary. Finally, relationally functional structure has been argued for many areas of language use, for example in adapting specific word orders in order to keep information flow uniform (Gibson et al. 2013).

Overenthusiastic functional explanations for linguistic rules are problematic, but this is equally the case for putting too much explanatory weight on functionally *unmotivated* rules. The problem seems clear: given that language is geared for social communication, why should any resources be allocated towards non-functional elements? This is used as a springboard into arguments about mechanism, innatenes, and so on. There are a few ways out of this: one option is that we have so far failed to identify the actual function. Another is that many rules are a mirage, a non-productive result of historical factors. Another alternative is a genuine lack of relevant function, despite rule-like behaviour having a real cause. This would imply that their function is epiphenomenal, a shadow of some other mental or physical process. Even this last option could be seen either as a purely linguistic phenomenon, albeit without function, as in Chomsky's conception of syntax: otherwise it could be the shadow of other mental or physical processes, along the lines of what Chomsky (2005) calls 'third factors'. A final analysis, however, is to recognise that function exists at different levels of analysis, and resides in different places. Syllable structure aids in communication in all its many formal variations: its function is structural without being tied to a specific structure.

To sum up, despite controversy over the amount and role of function in language, most of this is spill-over from related debates over mechanism, development, and

evolution. When function is identified, however, it is almost always social and communicative: in this case, it directly informs those debates. When function is unclear, however, it adds little, and care should be taken before invoking the *lack* of function as explanatory. Finally, it is important to acknowledge that function can be more or less specific, and more or less tied to particular structural features.

The acquisition of rules

Linguistic rules are highly idiosyncratic and variable, both within and across languages. All of these rules must be learned: the question is what needs to be learned, and how that takes place. This is a familiar topic in linguistics, where the *poverty of the stimulus* is often invoked to argue in favour of innate knowledge (e.g. Berwick et al. 2011). This usually targets syntactic knowledge, and not the enormous variety of rule-like behaviour found throughout the rest of language. Nevertheless, there are strongly nativist arguments found in some non-syntactic domains, such as phonology (e.g. Hale and Reiss 2008). The empirical and theoretical literature here is vast, so I will only sketch out some of the nativist positions as applied specifically to linguistic rules, along with some of their critiques.

Berent (2013), arguing for an innate phonological component, outlines 'seven wonders of phonology', including *early onset* and *regenesis*, i.e. that (1) humans acquire phonology 'rapidly and spontaneously', and that (2) phonology has reappeared *de novo* in recently emerged languages, citing a novel sign language (Al-Sayyid Bedouin Sign Language; Sandler et al. 2011) which appears to be in the process of recapitulating a phonological system. In support of the former, Berent cites results from her own work (Gervain et al. 2012) which suggests that neonates can learn phonological rules. However, this employs optical imaging studies of the brain to detect reliable correlates of rule learning, which have been criticised on technical and methodological grounds: Aslin (2012) argues that these results "fall short of answering the questions that motivated their authors... surely there is some region of cortex that responds to just about any stimulus" (p. 24)—although Aslinet al. (2015) cite the same work more positively. The results from sign language are also problematic, as Sandler et al. (2011, p. 536) actually conclude that ABSL "does not yet have a phonological level of structure", although "the kernels of a phonological system are beginning to present themselves". More specifically, the authors are referring to evidence that combinatorial phonological units *may* be emerging, but certainly nothing along the lines of phonological rules. In any case, Everett (2016) points out that even if Berent's empirical data is solid, this doesn't imply that children require an innate phonological component to learn phonological rules: even if there *is* an innate component, work needs to be done to differentiate a full-blown 'module' from a much less substantive learning bias, especially given children's well-established ability to perform statistical learning in the linguistic domain [for an overview, see Wonnacott (2013)].

Hale and Reiss (2008) employ formal learning theory to argue that both phonological units and principles are unlearnable *in principle* without access to innate knowledge. This claim centres around the 'Innateness of Primitives Principle' (p.

27), i.e. that a computational theory of learning always and only involves novel arrangements of pre-existing elements. They also outline a version of the *Subset Problem*: that to acquire a grammar that is not overly permissive, a learner must always choose the most restrictive subset of possible grammars. In Hale and Reiss's formulation, this requires that the initial, pre-learning state is *necessarily* richer than the subsequent one, thus implying the existence of rich innate knowledge. This argument, however, is predicated on what Hale and Reiss acknowledge (p. 39) is a fundamentally rationalist take on cognition, and ignores, for example, exemplar-theoretic approaches which allow for the *construction* of categories via statistical learning (e.g. Walsh et al. 2010). Also pushing against their analysis is the utility of *negative evidence* (i.e. learners are given examples of incorrect forms): they are justified in pointing out that explicit negative evidence is an unlikely and ill-defined source of information for language learners, but this really only applies to *explicit* negative evidence. However, Bayesian statistical learning (e.g. Tenenbaum and Griffiths 2001) can leverage *implicit* negative evidence easily, as demonstrated specifically for statistical phontactic learning by Linzen and Donnell (2015).

As such, both empirical data and theory supporting nativism for linguistic rules are inconclusive, and extremely weighed down by theoretical disposition. While there is a tendency for both sides in this debate to caricature the other, nobody disputes that there is an important role for learning: nativists want to reduce the burden on the learner by fleshing out a rich, language-specific cognitive module; their 'empiricist' opponents want to enrich the account of domain-general learning as much as possible before doing so. The question is really: are general processes of learning sufficient to learn linguistic rules? If not, how much do we need to ascribe to an innate linguistic module? Given what we now know about the power of human domain-general learning capacities, this seem to provide the most compelling solution: however, this is not to say that domain-specific adaptations play no role, but rather one in the form of weaker biases, as we shall see in the next section.

The evolution of rules

Human languages have taken shape via historical processes of transmission and change, but they would not exist without living, breathing, thinking humans to speak and transform them. As we have seen, some people would prefer to characterise linguistic rules as a primarily cultural phenomenon (e.g. Croft 2013), and others as primarily cognitive (e.g. Pinker and Bloom 1990). The former perspective seems to require that the rules emerge via cultural evolution alone, the latter the biological evolution of a dedicated module for handling linguistic rules. Of course, this apparent dichotomy need not be the case. Gene-culture co-evolution is uncontroversial in a number of non-linguistic cases (see Boyd and Richerson 2005): it is quite possible for both cognitive and cultural elements to be at play. Taking this view, the best question becomes: how feasible does the evolution of linguistic rules look under both biological and cultural evolutionary accounts?

Pinker and Bloom (1990) see the hand of natural selection acting on human biology as explaining their universals of human language. On their view, human

language is highly adapted for communication: "grammar is a complex mechanism tailored to the transmission of propositional structures through a serial interface" (p. 1). More recently, Pinker and Jackendoff (2005) argue for an innate phonological capacity. Theirs is a nativist position, but again not the only one: Berwick and Chomsky (2015), who take a resolutely nativist (and non-selectionist) position for syntax, are happy to embrace any number of domain-general, gene-culture co-evolutionary, or developmental explanations for the rest of language. But it is notable that many who take the strongest of nativist positions often feel no need to provide evolutionary explanation: Hale and Reiss (2008, p. 160) state that the innate faculty "must have come into being at some point", but that the past is "irrelevant" to synchronic concerns, and that, if phonology is an adaptation, then there is no reason why it should not be for the physiology of some much earlier form of hominid. When Everett (2016) criticises Berent (2013) for her failure to address the 'origins' question, she responds by acknowledging that "phonetic and sensorimotor pressures shape phonology in language evolution" (Berent 2016) but, similarly to Hale & Reiss, states that evolutionary history is not needed to establish the existence of a synchronic module. Running against these accounts, Blevins (2004, 2008) situates her *evolutionary phonology* entirely in the cultural sphere, without any recourse to a phonological module, let alone the evolution of one. What appears to be rule-like regularities is the result of incremental build-up of physiologically driven sound change, with analogical re-analysis gradually pulling more and more forms into a similar shape, such that it ultimately looks like there is a regular, productive rule. However, note that these are the exactly the *same* pressures which Berent (2016) suggests as shaping the *biological* underpinnings of phonology. These are not incompatible accounts—indeed, the Baldwin effect could plausibly be involved in the emergence of a phonological capacity, with rule-like behaviour developing first culturally, before subsequent genetic accommodation. But nobody involved in the phonological debate even mentions such an idea. This is to be expected: Hale and Reiss (2008) argue as hard for a purely synchronic take on linguistics as Blevins (2004) does for a purely diachronic one: the synchronic view assumes an innate component but is (at best) agnostic about its origin; the historical view makes do without an innate component and therefore finds no need for one. And besides, the historical take operates over very different timescales than required for genetic adaptation.

This feeds into a more general debate on the possibility of the Baldwin effect for language: Chater et al. (2009) address its possibility, using computer simulations to show that genetic accommodation of such a rapidly changing cultural object as language is highly unlikely. Note, however, that this is not the case if there are more stable features involved in language which evolution might be able to track. The physiological pressures cited by both sides are an excellent candidate for such a stable feature, but other stabilities might exist more in the form of general trends in the cultural development of rule-like structures. In this case, it seems reasonable to expect *some* genetic accommodation towards a set of general principles which shape linguistic rules. Whether this looks like the frameworks set out by formal linguists, or closer to a (possibly weak) *cognitive bias* (see Thompson et al. 2016) is a matter of mechanism, not evolution, and this is where cognitivists must convincingly argue

their case. But, as both sides acknowledge the same guiding pressures, and at least the *existence* of longer time-scales over which language has been spoken, the onus of explanation is also on non-nativists to explain why either there was no genetic adaptation, or that it formed in a way irrelevant to linguistic rules.

A minimal account of linguistic rules

We can now use Tinbergen's framework to map out the main positions:

1. *Mechanism*
   - sophisticated, domain (and sub-domain) specific
   - sophisticated, domain-general mechanisms for processing and interaction
2. *Function*
   - externally but not ultimately for social communication because:
     - no ultimate function
     - ultimately non-communicative function
   - ultimately for social communication
3. *Acquisition*
   - learning selects a subset of rules from rich, innately-specified knowledge
   - learning infers rules from rich, culturally-transmitted knowledge
4. *Evolution*
   - biological evolution alone
   - cultural evolution alone
   - gene-culture co-evolution

I will now outline what I regard as the *minimal* position in light of the evidence, including empirical knowledge about both non-human and non-linguistic human cognition. This is that linguistic rules are (1) mainly—but not necessarily entirely—governed by sophisticated domain-general mechanisms, (2) have a general but not necessarily specific communicative function; (3) are inferred from culturally transmitted knowledge with aid from certain cognitive and physiological biases, and (4) evolve mainly culturally, with limited genetic accommodation.

Starting with mechanism, there is evidence that primates neurologically encode sophisticated behaviour, with stimulation of specific brain sites eliciting "complex, apparently meaningful movements" (Graziano 2006, p. 120). Alongside evidence for evolutionarily novel and functionally redeployed networks in the human brain (Mantini et al. 2013), this suggests that there is no *a priori* reason why linguistic

rules should not be biologically 'encoded in the brain'.[5] However, it is unclear what exactly *is* encoded in even these relatively simple cases: take, for example, arm-reaching behaviour: "there remains no consensus about what motor cortex is specifying about the reach. It remains unclear whether cells represent a signal for muscle force, the direction of movement, or a more abstract end goal of muscle action" (Lisman 2015, p. 874). More importantly, Wallis et al. (2001) show that monkeys can be trained to learn abstract rules which are *also* encoded by a single neuron: being neurologically encoded has no necessary implications regarding questions of innateness or specificity. Indeed, Christiansen and Chater (2016) argue that very *general* demands on processing create a 'now or never bottleneck': these require powerful, domain-general mechanisms for "eager processing, computing multiple representational levels, and anticipation" (p. 4), and it is these which are responsible for rule-like behaviour in language. Notably, it is the link to linguistic form which is controversial: nobody denies the sophistication of human processing capacities. This is important: given that both non-linguistic and linguistic processing are undeniably sophisticated, why should we assume that the mechanism responsible for linguistic rules is completely unrelated? This, then, should be the point of departure: to show that this kind of mechanism is or is not capable of explaining all rule-like behaviour in language. In fact, even proponents of syntactic nativism have argued for this kind of explanation for most external aspects of language (e.g. Samuels 2009; Berwick and Chomsky 2015).

Linguistic rules have a social, communicative function in a broad, almost trivial sense: the question is the extent to which any particular form lends itself to that function. Some of the more general universals of form (Pinker and Bloom 1990; Berent 2013) arguably speak more to shared function than, as intended, mechanism. These are properties shared by many quite differently instantiated rule systems, and ones with reasonably overt function: for example, syllables and vowels break up the speech stream. However, there are many different ways to build a syllable, so this may be (and in reality is) largely unconstrained. Many architectural principles in biology are like this. Snail shells share many features, such as chirality, but the direction of coiling is presumably neutral to function, beyond considerations of reproductive compatibility (although even this does not always apply, see Schilthuizen and Davison 2005). While there is a general protective function, many of the details are functionally under-determined. Where there are claims for specific rules having form-specific functional motivation, this needs to be cashed out empirically, often no easy task: investigating explanations for adjective order in English, Wulff (2003) found that the best fit when employing a mixture of *all* variables proposed in the literature still gave only 73% empirical coverage. However, it is worth stressing once again that the failure to identify function at a particular level of analysis does not have any logical consequences for mechanism, acquisition, or evolution: function may reside at some other level.

[5] However, as an anonymous reviewer points out, there is an extensive literature on potential similarities between action sequencing and language processing which has yet to resolve itself into anything resembling a consensus, so we should remain cautious about making any categorical statements here.

Monkeys can learn abstract rules, if not particularly complex ones (Wallis et al. 2001; Fitch 2004); humans and non-human primates share neural circuitry associated with rule learning (Bunge 2005); seven-month-old human infants can learn 'algebriac' rules in an artificial language task (Marcus 1999), but this kind of rule learning is not specific to language (Saffran et al. 2007). Despite set-theoretic arguments for the logical necessity of innate knowledge, general learning mechanisms are able to learn rule-like structures, whether via distributed/neural-network (e.g. Elman 1991) or Bayesian (e.g. Griffiths et al. 2007) methodologies. Given the preponderance of evidence that not only is rule learning easily handled by general mechanisms, but also evolutionarily well-established, it is hard to argue for a strong role for innate knowledge in the acquisition of linguistic rules. On the other hand, innate knowledge in the form of a weaker learning bias is quite possible (e.g. White 2014).

The evolutionary story favoured by this largely domain-general, communicatively-driven, learning-heavy account is a mainly cultural one. Linguistic rules take shape over the course of multiple repeated instances of use, learning, and replication; individual linguistic items share increasingly more structural properties as analogical processes intervene. This resembles something like the naturalist take on phonology expressed by Blevins (2004), but is not necessarily so hard-line. However, this must be assessed alongside an growing body of work which suggests that, when genes and culture co-evolve, some genetic accommodation is likely in the form of a weak, defeasible bias (e.g. Thompson et al. 2016).

A final mention should be made of one factor which ties up mechanism, function, learning and evolution: a number of researchers have proposed similar arguments that, given language is a culturally acquired system processed by human cognition, it is likely that language culturally evolves not just for its communicative function, but also for its 'cognitive niche', i.e. so that we can learn and process it optimally (e.g. Christiansen and Chater 2008; Kirby et al. 2015). This is a compelling account, and one that fits well into Tinbergen's framework. An implication this has for linguistic rules, which has not yet been covered, relates to learnability. Rule-like structures have the added benefit of being *predictable* and *regular*. This provides a scaffold for learning new forms which conform to this structure, and if that general structure has communicative function, should help that rule persist and spread throughout the linguistic system.

Taken together, we can give a reasonably deflationary account of linguistic rules. Moreover, it is an *extensible* one. Few would argue against the existence of general learning capacities in humans, or that language is not a culturally evolving communicative tool. But this still leaves open the possibility that certain design features—and conceivably even whole domains—of rule-like behaviour in language can not be explained under this framework. However, the burden of proof lies on those who want to show that, for example, some rule-producing mechanism is language-specific, has no communicative function, or is an innate biological component with no evolutionary history. Nevertheless, there are a few ways in which people might push against this story.

One likely objection to the account presented here is the existence of some linguistic phenomenon requiring a mechanism which is either more sophisticated or

more plausible than the general learning capabilities proposed here. If our topic were syntax, this would represent a more formidable problem. However, recent empirical approaches to language learnability (Chater et al. 2015) have shown that general, distributional approaches are capable of learning even complex syntactic structures, and the complexity of non-syntactic linguistic rules is generally assumed to be far less (e.g. Kaplan and Kay 1994). On a different tack, the human ability to parse and use symbolic rules compared to other primates might be seen as a matter of measure: we do so promiscuously, other primates only under duress and experimental conditions. From this perspective, primates possess the building blocks, but humans have composed them. This is a reasonable assumption, but one which can be made about a whole slew of human cognitive capacities, especially as relate to social functions: it does not speak to claims of innateness or domain specificity. Finally, another line of argument is that it is impossible to treat linguistic rules in the general, without reference to specific but large disciplinary sub-domains such as phonology and morphology, each requiring its own explanation. Indeed, it is certain that these areas do apply their own peculiar concerns, to particular effect: my argument is that we can rely on a minimal account for the vast majority of cases of linguistic rule without recourse to such domain-specific mechanisms.

We can return to the motivating question: what are linguistic rules, why are they so ubiquitous, and why do they share some features while also displaying such variation? The answer I have been arguing for is this: linguistic rules are a culturally evolving feature of language. While they may not represent a domain-specific, innately-specified cognitive module, they do involve sophisticated cognition, tying together processes of fast processing and analogy with a rich and expansive memory. They are pervasive because they share a broadly communicative function, which spreads via analogy: they display such variation because this function is not tied to a particular form, but is multiply realisable.

References

Anderson SR (1981) Why phonology isn't "natural". Linguist Inq 12(4):493–539
Aslin RN (2012) Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy. Cogn Neuropsychol 29(1–2):7–33
Aslin RN, Shukla M, Emberson LL (2015) Hemodynamic correlates of cognition in human infants. Annu Rev Psychol 66(1):349–379
Berent I (2013) The phonological mind. Trends Cogn Sci 17(7):319–327
Berent I (2016) Commentary: "an evaluation of universal grammar and the phonological mind"—UG is still a viable hypothesis. Front Psychol 7(JUL):1–13
Berwick RC, Chomsky N (2015) Why only us?. MIT Press, Cambridge
Berwick RC, Pietroski P, Yankama B, Chomsky N (2011) Poverty of the stimulus revisited. Cogn Sci 35(7):1207–1242
Blevins J (2004) Evolutionary phonology: the emergence of sound patterns. Cambridge University Press
Blevins J (2008) Natural and unnatural sound patterns: a pocket field guide. In: Willems K, De Cuypere L (eds) Naturalness and iconicity in language. John Benjamins, Amsterdam, pp 121–148
Boeckx C (2010) Language in cognition. Wiley-Blackwell, Hoboken
Boyd R, Richerson PJ (2005) The origin and evolution of cultures, vol 43, No. 02. Oxford University Press, Oxford
Brown K (ed) (2005) Encyclopedia of language and linguistics. Elsevier Science, Amsterdam

Buckley E (2000) What should phonology explain? Handout from SUNY Buffalo linguistics colloquium.
Bunge SA (2005) Neural circuitry underlying rule use in humans and nonhuman primates. J Neurosci 25(45):10347–10350
Chater N, Clark A, Goldsmith JA, Perfors A (2015) Empiricism and language learnability. Oxford University Press, Oxford
Chater N, Reali F, Christiansen MH (2009) Restrictions on biological adaptation in language evolution. Proc Natl Acad Sci USA 106(4):1015–1020
Chomsky N (2005) Three factors in language design. Linguist Inq 36(1):1–22.
Chomsky N, Halle M (1968) The sound pattern of English. Harper, New York
Christiansen M, Reali F, Chater N (2011) Biological adaptations for functional features of language in the face of cultural evolution. Hum Biol 83(2):247–259
Christiansen MH, Chater N (2008) Language as shaped by the brain. Behav Brain Sci 31(5):489–508 discussion 509–558
Christiansen MH, Chater N (2016) The Now-or-Never bottleneck: a fundamental constraint on language. Behav Brain Sci 39:1–52
Croft W (2001) Radical construction grammar: syntactic theory in typological perspective. Oxford University Press, Oxford
Croft W (2013) Evolution: language use and the evolution of languages. In: Binder PM, Smith K (eds) The language phenomenon. The frontiers collection. Springer, Berlin, pp 93–120
de Lacy P, Kingston J (2013) Synchronic explanation. Nat Lang Linguist Theory 31(2):287–355
Elman JL (1991) Distributed representations, simple recurrent networks, and grammatical structure. Mach Learn 7(2):195–225
Everett DL (2016) An evaluation of universal grammar and the phonological mind. Front Psychol 7(15):1–10
Fitch WT (2004) Computational constraints on syntactic processing in a nonhuman primate. Science 303(5656):377–380
Gervain J, Berent I, Werker JF (2012) Binding at birth: the newborn brain detects identity relations and sequential position in speech. J Cogn Neurosci 24(3):564–74
Gibson E, Piantadosi ST, Brink K, Bergen L, Lim E, Saxe R (2013) A noisy-channel account of crosslinguistic word-order variation. Psychol Sci 24(7):1079–88
Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc B Biol Sci 205(1161):581–598
Graziano M (2006) The organization of behavioral repertoire in motor cortex. Annu Rev Neurosci 29(1):105–134
Griffiths TL, Chater N, Kemp C, Perfors A, Tenenbaum JB (2010) Probabilistic models of cognition: exploring representations and inductive biases. Trends Cogn Sci 14(8):357–364
Griffiths TL, Kemp C, Tenenbaum JB (2007) Bayesian models of cognition, pp 1–49
Griffiths TL, Lieder F, Goodman ND (2015) Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. Top Cogn Sci 7(2):217–229
Hale M, Reiss C (2008) The phonological enterprise. Oxford University Press, Oxford
Huddlestone R, Pullum GK (2002) The Cambridge grammar of the English language. Cambridge University Press, Cambridge
Johansson S (2013) Biolinguistics or physicolinguistics? Is the third factor helpful or harmful in explaining language? Biolinguistics 2005:249–275
Kaplan RM, Kay M (1994) Regular models of phonological rule systems. Comput Linguist 20:331–378
Kirby S, Tamariz M, Cornish H, Smith K (2015) Compression and communication in the cultural evolution of linguistic structure. Cognition 141:87–102
Linzen T, Donnell TJO (2015) A model of rapid phonotactic generalization. In: Emnlp, pp 1126–1131
Lisman J (2015) The challenge of understanding the brain: where we stand in 2015. Neuron 86(4):864–882
Mantini D, Corbetta M, Romani GL, Orban GA, Vanduffel W (2013) Evolutionarily novel functional networks in the human brain? J Neurosci 33(8):3259–3275
Marcus GF (1999) Rule learning by seven-month-old infants. Science 283(5398):77–80
Marr D, Poggio T (1976) From understanding computation to understanding neural circuitry
Ohala JJ (2005) Phonetic explanations for sound patterns: implications for grammars of competence. In: A figure of speech. A festschrift for John Laver, pp 23–38
Oudeyer P-Y (2005) The self-organization of combinatoriality and phonotactics in vocalization systems. Connect Sci 17(3–4):325–341

Pinker S (1991) Rules of language. Science 253(5019):530
Pinker S (1999) Words and rules. Basic Books Inc, New York
Pinker S, Bloom P (1990) Natural language and natural selection. Behav Brain Sci 13(13):707–784
Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? Cognition 95(2):201–36
Saffran JR, Pollak SD, Seibel RL, Shkolnik A (2007) Dog is a dog is a dog: infant rule learning is not specific to language. Cognition 105(3):669–680
Samuels B (2009) The third factor in phonology. Biolinguistics 3(2–3):355–382
Sandler W, Aronoff M, Meir I, Padden C (2011) The gradual emergence of phonological form in a new language. Nat Lang Linguist Theory 29(2):503–543
Schilthuizen M, Davison A (2005) The convoluted evolution of snail chirality. Naturwissenschaften 92(11):504–515
Strunk W, White EB (1979) The elements of style. Allyn & Bacon, Boston
Tenenbaum JB, Griffiths TL (2001) Generalization, similarity, and Bayesian inference. Behav Brain Sci 24(04):629–640
Thompson B, Kirby S, Smith K (2016) Culture shapes the evolution of cognition. Proc Natl Acad Sci USA 113(16):4530–4535
Tinbergen N (1963) On aims and methods of ethology. Z Tierpsychol 20:410–433
Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. Nature 411(6840):953–956
Walsh M, Möbius B, Wade T, Schütze H (2010) Multilevel exemplar theory. Cogn Sci 34(4):537–582
White J (2014) Evidence for a learning bias against saltatory phonological alternations. Cognition 130(1):96–115
Wonnacott E (2013) Learning: statistical mechanisms in language acquisition. In: Binder PM, Smith K (eds) The language phenomenon. The frontiers collection. Springer, Berlin, pp 65–92
Wulff S (2003) A multifactorial corpus analysis of adjective order in English. Int J Corpus Linguist 8(2):245–282
Zuidema W, de Boer B (2009) The evolution of combinatorial phonology. J Phon 37(2):125–144